



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE CORRELATION BETWEEN LENGTH OF FLOWERING-STALK
AND NUMBER OF FLOWERS PER INFLORESCENCE IN
NOTHOSCORDUM AND ALLIUM.

BY J. ARTHUR HARRIS.

Several years ago DeVries in outlining some of the problems of chief importance for future investigation in evolutionary work, designated correlation as one of the most important.* In another place in the same volume he discusses data of his own for the relationship between the diameter of the stem and the length of the fruit in certain of his evening primroses.†

Notwithstanding DeVries's emphasis of the importance of quantitative studies of correlation, comparatively little has been done by botanists. His suggestion concerning a relationship between the size of the individual and the measurable characters of the fruit called my attention to the desirability of a series of comparative studies of the interrelationship between the vegetative development of organs of the individual and the number or fertility of the flowers or fruits which it produces.

As material for a first determination of the interdependence of length of flowering-stalk and number of flowers per inflorescence I shall use series of *Nothoscordum* and *Allium*. Data from the inflorescences of several other species are being examined.

In the spring of 1906, a collection of the flowering-stalks of *Nothoscordum striatum* was made just south of the railway station at Meramec Highlands, near St. Louis, Mo. Here the brow of the hill slopes in two directions, to the southeast and to the southwest. The ground was more or less covered with old dead or leafless vegetation consisting chiefly of sumach bushes, large tufts of grass and weeds, and low grass. While no quantitative determination was made, it seemed to me that the old vegetation on the southwest slope was much

* DeVries, H. Die Mutationstheorie. 1:113. 1901.

† DeVries, l. c. 381-383.

more diversified from quadrat to quadrat in size and quantity than that upon the southeast slope. These remarks apply merely to the localities from which the samples of material were taken.

It occurred to me that a consequence of this condition might be a greater variability of the length of the stem in the series from the southwest slope. Here some of the plants grew up through large tufts of grass or dead weeds, while others occurred in open spots. Some were close to shrubs of considerable size and others grew where the vegetation was much smaller. On the southeast slope, on the other hand, the grass was rather uniformly short and weeds and bushes very few. If variability in the individuals of a population be to a considerable extent due to diversity in their edaphic environment, as many writers have suggested, it would seem quite reasonable to expect a greater variability on the southwest slope where conditions were apparently more diverse than on the southeast slope.

TABLE I. NOTHOSCORDUM.

		Length of Stalk in Centimeters.																																		
		9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	Totals								
Flowers per Umbel	3	1	1	3	1	3						
	4	..	1	1	3	4	2	11	6	3	1	3	2	1	2	1	1	42						
	5	2	..	3	2	2	10	11	11	6	15	8	4	2	3	6	1	3	89						
	6	..	1	..	1	1	8	2	7	8	10	8	10	9	7	3	3	2	1	1	82						
	7	1	1	1	2	3	2	2	7	6	11	9	8	13	4	3	..	2	1	1	1	1	1	..	79							
	8	2	5	2	2	5	2	2	6	3	..	2	1	2	..	2	4	29							
	9	1	1	1	1	1	1	..	1	..	4	4	2	1	2	1	1	2	2	24							
	10	1	1	2	2	6							
	11	1	1	1	1	1	..	5						
	Totals	2	2	6	7	9	23	28	28	19	36	30	31	27	25	31	13	8	6	8	7	4	3	..	4	1	1	..	359							

CORRELATION IN FIRST COLLECTION, SOUTHWEST SLOPE.

For the purpose of ascertaining whether the plants from these two spots, distant only a few yards from each other, do show differences in variability and correlation the material was collected in separate lots. That from the southwest slope comprised 359 flowering-stalks and the correlation be-

tically all the flowers had opened or were nearly ready for anthesis. All those with young buds were discarded to avoid the possibility of including stalks which had not completed their growth. These were very few in number and the samples taken may be regarded as representative of the species in this habitat. The rather coarse scale of centimeter units was adopted partly because of the possible error in the collecting of the material referred to above and partly to obviate the apparently needless labor of dealing with much larger correlation tables.

TABLE A. NOTHOSCORDUM.

Organ and Source of Material	Average and Probable Error	Standard Deviation and Probable Error	Coefficient of Variation
Length of Stalk			
Southwest Slope	$19.407 \pm .162$	$4.545 \pm .114$	23.421
Southeast Slope	$17.970 \pm .115$	$3.511 \pm .081$	19.541
Difference	$+1.437 \pm .198$	$+1.034 \pm .140$	$+3.880$
Flowers per Inflorescence			
Southwest Slope	$6.212 \pm .056$	$1.584 \pm .040$	25.505
Southeast Slope	$5.423 \pm .039$	$1.217 \pm .028$	22.435
Difference	$+.789 \pm .068$	$+.367 \pm .049$	$+3.070$

VARIATION CONSTANTS.

The means, standard deviations and coefficients of variation, and the probable errors of the first two constants, are given for length of stalk and number of flowers per inflorescence in Table A. Sheppard's correction for the second moment was used for all the variabilities given in this paper.

These constants require little discussion in this place. The means, standard deviations and coefficients of variation are all somewhat higher, and as the probable errors of the differences show significantly higher, in the series from the southwest slope than in the other of the two lots. The empirical range is also wider, although N is smaller, in the collection from the southwest slope, where I had estimated that edaphic conditions were more variable than in that from the southeast slope where they had appeared to me more uniform.

But before any great stress is laid upon these results some real measure of the relative variability of the environmental conditions in different habitats should be secured. Until better data can be obtained these differences in the constants for two collections of material in habitats a stone's throw from each other may serve as a further warning against the collection of biometric material without due regard to environmental conditions.*

The coefficients of correlation are the constants which give the direct answer to the chief question in hand. They are:—

Plants from southwest slope, $r = .549 \pm .025$

Plants from southeast slope, $r = .490 \pm .025$

Difference..... $.059 \pm .035$

These constants show at once that there is a very substantial interdependence between the length of the flowering-stalk and the number of flowers per inflorescence. The two constants appear somewhat unlike to the eye, but when the probable error of their difference is taken into account it is clear that their divergence is not significant.

To render the relationship between the length of the stem and the number of flowers per inflorescence more clear to those unacquainted with biometric terminology we may express the relationship in terms of regression instead of correlation. The equations to the regression straight lines are:—

For the series from the southwest slope,

$$y = 2.501 + .191 x.$$

For the series from the southeast slope,

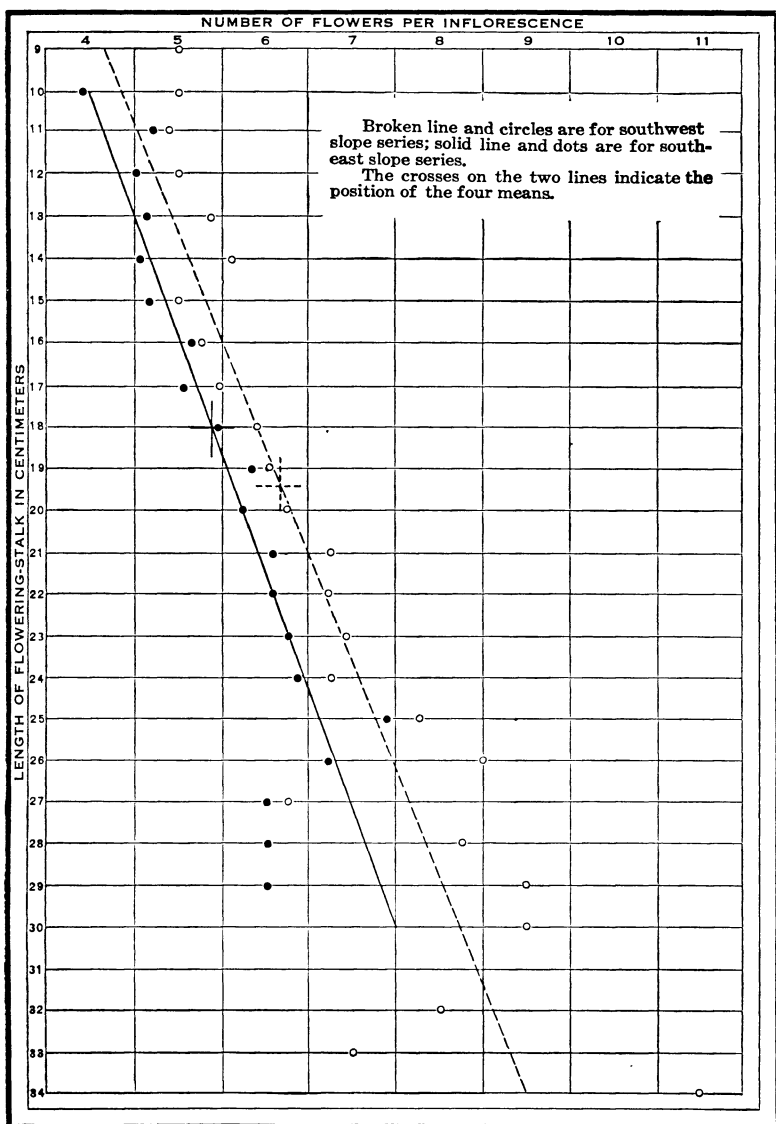
$$y = 2.375 + .170 x.$$

Here x denotes length of stalk and y number of flowers. From these regression coefficients we see that an increase of about two-tenths of a flower per inflorescence is on an average associated with an increase of a centimeter in the length of the stalk.

The relationship between the length of the stalk and the number of flowers per inflorescence may be represented graphically by the slope of the regression line, as in Diagram I.

* For one of the first discussions of this subject see the papers on the sources of apparent polymorphism in plants in *Biometrika*, **1**: 304–319. 1902.

DIAGRAM I. NOTHOSCORDUM.



SLOPE OF REGRESSION STRAIGHT LINES IN TWO COLLECTIONS.

The solid dots and the circles show the actually observed mean numbers of flowers per inflorescence for the several grades of stem length in the two series. The agreement between observation and theory is not very close, but N is not very large in either series and consequently the means of some of the arrays are based upon an inadequate number of observations. While the observed means do not fall so near the theoretical lines as one might like to see them, it seems quite idle to seek for a better equation than that for the straight line. Regression seems to be, so far as we can judge from the present data, linear.

The material for *Allium stellatum* was taken in the fall of 1907 from the hills at Meramec Highlands, Mo. The plants were not collected from such restricted habitats as those of *Nothoscordum*, but were drawn from a considerable range of territory, perhaps four square miles altogether. The method of gathering and measuring the stalks was the same as for *Nothoscordum*. In this case the range of variation is so wide that it is convenient to group in classes of three units for purposes of calculation. The data are presented in Table III; the calculated variation constants, in Table B. The coefficient of correlation between length of stalk and number of flowers per inflorescence is $.619 \pm .018$. Here again the variation constants require no particular comment and are given merely for the convenience of the reader. The coefficient of correlation is slightly higher than that for *Nothoscordum*, but whether any significance is to be attached to the difference could only be determined by the collection of further series of material.

The equation to the regression straight line is

$$y = -15.480 + 1.311 x,$$

where x = length of stalk and y = number of flowers per inflorescence. The graph, Diagram II, shows that the empirical means fall very close to the theoretical line for all but the last three classes. Here there appears at first to be some indication of non-linear regression, but a glance at the correlation table shows that the means are based on only 4, 3, and 1 observations respectively and hence cannot be given

much weight. Under these circumstances I would rather devote time to the collection of additional data than to the mathematical analysis of those already on hand. Hence no further tests for linearity of regression are applied.

TABLE III. ALLIUM.

		Number of Flowers per Inflorescence.																																		
		8-10	11-13	14-16	17-19	20-22	23-25	26-28	29-31	32-34	35-37	38-40	41-43	44-46	47-49	50-52	53-55	56-58	59-61	62-64	65-67	68-70	71-73	74-76	77-79	80-82	83-85	86-88	89-91	92-94	95-98	99-101	102-104	105-107	108-110	
Length of Stalk in Centimeters	26-28	1	..	1	1	..	2	1	..	3	..	2	..	1	5
	29-31	2	5	..	6	15	10	9	9	5	19	
	32-34	2	..	4	2	..	7	8	14	21	15	16	4	5	3	73	
	35-37	106	
	38-40	..	1	..	2	..	3	9	13	11	16	14	23	10	5	5	4	6	6	2	120	
	41-43	3	3	3	7	11	17	10	9	10	2	4	4	3	3	82	
	44-46	2	1	1	1	2	9	8	7	10	6	4	4	4	4	1	1	1	64	
	47-49	2	1	1	3	1	8	8	2	2	1	2	2	1	1	1	32	
	50-52	2	..	2	2	2	1	2	1	3	3	1	1	1	16	
	53-55	4	
	56-58	2	3	
	59-61	1
		3	1	7	14	20	40	42	55	65	61	57	40	37	21	23	13	6	8	1	4	2	1	1	1	..	1	..	1	..	1	1	525

CORRELATION.

TABLE B. ALLIUM.

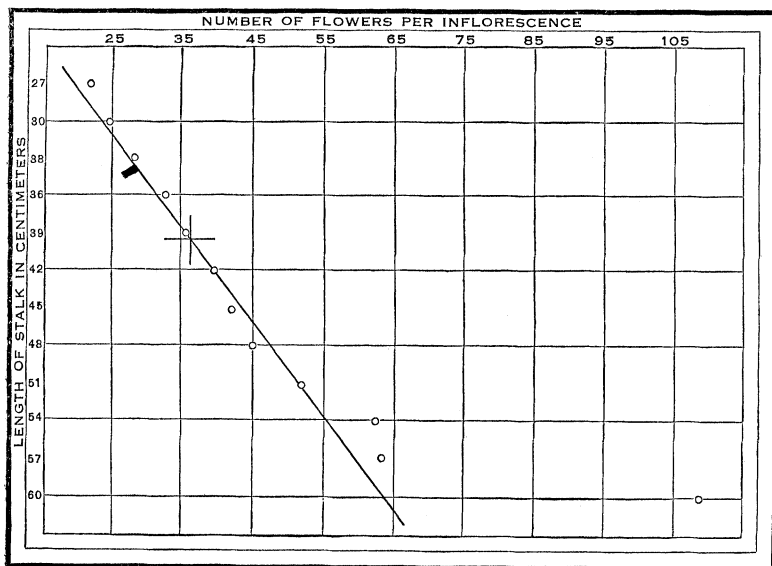
	Length of Stalk	Number of Flowers per Inflorescence
Average and Probable Error	39.491 ± .160	36.309 ± .340
Standard Deviation and Probable Error	5.446 ± .118	11.571 ± .249
Coefficient of Variation	13.791	31.785

VARIATION CONSTANTS.

In both of these cases the regression seems to be linear, although in *Nothoscordum* the fit of the means to the theoretical line is not very good, and in the case of *Allium* there is some question concerning the upper part of the range, where the observations are inadequate. Further tests of linearity might be applied, but it seems hardly worth while to do so at the present time.

The significance of linear regression in the present cases is two-fold. First, it shows that we are quite justified in using the coefficient of correlation to describe the degree of interdependence of length of flowering stem and number of flowers per inflorescence. Second, it shows our material to be in

DIAGRAM II. ALLIUM.



SLOPE OF REGRESSION STRAIGHT LINE.

this respect in agreement with the vast majority of organic relationships, for in most cases of organic correlation regression is sensibly linear.

In short the high positive values of r and the apparently linear regressions indicate that in this material the length of the stalk and the number of flowers per umbel are intimately associated and that the rate of change in the number of flowers as we pass from the shortest to the longest stalks remains practically the same throughout the entire range of variation of the latter character.

There are sources of danger in interpreting these coefficients of correlation biologically. We are dealing here with statistical constants which measure precisely the degree of similarity of the associated variables under consideration, but give no intimation whatever as to the cause of this similarity. The physiologist has long used the term correlation, and it must be borne in mind that the word has a somewhat different meaning in the terminology of quantitative biology. All that the coefficient of correlation does is to *describe*, not *interpret*, degrees of interdependence. In the present instance we are dealing with perennial plants, and it is not at all unreasonable to think that the age of the individuals may make considerable difference in the magnitude of both the length of the flowering-stalk and the number of flowers per inflorescence. In this case the correlation would be increased by an amount depending upon the age heterogeneity of the collections. There seems to be, however, no way of avoiding this difficulty in material of this kind gathered in nature. Comparison with plants of annual habit of growth might, perhaps, throw some light upon this question.

In conclusion, the data presented show that there is a considerable degree of interdependence between the length of the flowering-stalk of these monocotyledons and the number of flowers which they produce. It is not at all surprising to find this relationship, and refined statistical methods would not have been required to demonstrate its existence in this particular case. To those who urge that it is pedantic to devote laborious researches to the demonstration of what is to be *a priori* expected, the reply is two-fold. First, in science, assumptions, however plausible on the surface, should never be accepted without actual demonstration on adequate material. The need of careful regard to this dictum is found in the group of problems we are considering here. In unpublished data somewhat similar to these, for other species, I find results not at all in agreement with those set forth in the present paper, although there is *a priori* quite as much reason for expecting high positive correlations as in the species here considered. Second, we should not be content to merely demonstrate the existence of a relationship; we

should measure its intensity as well. When a sufficient number of such measurements are available, comparisons will permit broader generalizations and suggest deeper researches than we can now foresee, or can ever hope to attain without strictly quantitative methods of investigation.